Surface Water and Ocean Topography (SWOT) Project

SWOT Product Description
Long Name: Level 2 KaRIn high rate lake single pass vector product
Short Name: L2_HR_LakeSP

Revision A (DRAFT)

Prepared by:

Cassie Stuurman
JPL Algorithm Engineer

Claire Pottier
CNES Algorithm Engineer

Approved by:

Curtis Chen
JPL Algorithm System Engineer

Roger Fjørtoft
CNES Algorithm System Engineer

Concurred by:

Oh-Ig Kwoun
JPL SDS Manager

Hélène Vadon
CNES SDS Manager

Paper copies of this document may not be current and should not be relied on for official purposes. The current version is in the JPL Engineering Product Data Management System (EPDM: https://epdm.jpl.nasa.gov) and the CNES Product Data Management System

June 4, 2020
SWOT-TN-CDM-0673-CNES

National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology

CHANGE LOG

VERSION	DATE	SECTIONS CHANGED	REASON FOR CHANGE
Preliminary	2017-04-26	All	Preliminary version
Initial Release	2019-09-20	All	Initial Release
Initial Release V2	2020-03-23	All	Updates following SME review
Revision A	2020-mm-dd	All	Product divided into 3 shapefiles

Table of Contents

Table of Figures

Table of Tables

List of TBC Items

Page	Section
10	§3.2: Format of PLD-oriented lake file: shapefile? (or other format without polygons?)
10	§3.2: The L2_HR_LakeSP product contains a record for each observed lake, greater than 1 ha
20	§4.3: Size of lakes above which the standard deviation of the height is computed (currently 5000ha)

List of TBD Items

Page	Section
9	$2.1: Reference to User handbook for HR mask to be added when ready.
17	§4.1.1: Number of digits in the identifiers (o*bs_id* and *lake_id*)
22, 33	§4.1.3, 5.2: Content of *layovr_val* parameter (metric of layover effect)
23, 35	§4.1.5, 5.2: Content of *xovr_cal_q* (quality of the cross-over calibration)
40	§6: Reference of “SWOT L2_HR Prior River Database Auxiliary File Description”
44	§Appendix B: XML file content.

1 Introduction

1.1 Purpose

The purpose of this Product Description Document is to describe the Level 2 Ka-band Radar Interferometer (KaRIn) high rate (HR) lake single pass (SP) vector data product from the Surface Water Ocean Topography (SWOT) mission. This data product is also referenced by the short name L2_HR_LakeSP.

1.2 Document Organization

Section 2 provides a general description of the product, including its purpose and latency.

Section 3 provides the structure of the product, including granule definition, file organization, spatial resolution, temporal and spatial organization of the content, file size, and overall data volume.

Section 4 provides qualitative descriptions of the the information provided in the product.

Section 5 provides a detailed identification of the individual fields within the L2_HR_LakeSP product.

Section 6 provides references.

Appendix A provides a list of the acronyms used in this document.

Appendix B provides a description of the format of the product metadata.

1.3 Document Conventions

When the specific names of data variables and groups of the data product are given in the body text of this document, they are usually represented in *italicized text*.

2 Product Description

2.1 Purpose

The L2_HR_LakeSP product provides lake data from each continent-pass of the high-rate (HR) data stream of the SWOT KaRIn instrument. These data are generally produced for inland and coastal hydrology surfaces, as controlled by the reloadable KaRIn HR mask [TBD].

Rivers in the Prior River Database (PRD) [1] are included in the Level 2 KaRIn High Rate River Single Pass Vector Product (L2_HR_RiverSP) [2]. As discussed further in Section 3.2, the L2_HR_LakeSP product specifically provides data for lakes identified in the Prior Lake Database (PLD) [3], and for detected features that have not been identified as a lake in the PLD nor as a river in the PRD.

Note that lakes connected to a river topology in the PRD will be included in both L2_HR_LakeSP and L2_HR_RiverSP products.

2.2 Latency

The L2_HR_LakeSP product is generated with a latency of at most 45 days from data collection. The latency allows for consolidation of instrument calibration and the required auxiliary or ancillary data that are needed to generate this product. Different versions of the product may be generated at different latencies and/or through reprocessing with refined input data, such as an updated version of the PLD.

3 Product Structure

3.1 Granule Definition

The L2_HR_LakeSP product is provided in full-swath pass granules (i.e., including both left and right half swaths) covering individual continents, with continent boundaries defined by the PLD, as described in [1]. These continent boundaries are consistent with the associated Level 2 KaRIn High Rate River Single Pass Vector Product (L2_HR_RiverSP) [2]. The terms "left" and "right" are defined as if standing on the Earth surface at the spacecraft nadir point facing in the direction of the spacecraft velocity vector. The L2_HR_LakeSP granule covers a swath that is approximately 128 km wide in the cross-track direction, although SWOT performance requirements are only applicable from 10–60 km from nadir on each side; observations may be missing, degraded, and/or flagged over the central 20 km of the swath. A "pass" is a half revolution of the Earth by the satellite from pole to pole (south to north latitudes for ascending passes, and north to south latitudes for descending passes).

3.2 File Organization

The L2_HR_LakeSP product is distributed in the Esri Geographical Information System (GIS) shapefile format [4]. Each granule of the product consists of three shapefiles:

- an observation-oriented shapefile of lakes identified in the PLD
- a PLD-oriented shapefile (I) of lakes identified in the PLD
- a shapefile of unassigned features (i.e., not identified in PLD nor PRD)

The observation-oriented L2_HR_LakeSP lake shapefile contains one record for each observed lake (identified in the PLD) greater than 1 ha (TBC), that is covered by the granule. Note that one observed lake may correspond to several PLD lakes.

The PLD-oriented L2_HR_LakeSP lake shapefile likewise contains lakes greater than 1 ha (TBC), but with one record per PLD lake. Each record can therefore consist of (partial) polygons of several observed lakes.

The PLD L2_HR_LakeSP unassigned features shapefile is observation-oriented and includes records for detected water bodies that have neither been identified as a lake in the PLD, nor as a river in the PRD. As illustrated in Figure 1, the observation- and PLD-oriented lake shapefiles are redundant to a large extent, and provided to accommodate different user needs. The observed lakes may be:

- connected lakes, i.e. lakes that have either an inflow or an outflow or both in the SWOT river network, as defined in the PRD (referenced both in the PLD and the PRD).
- disconnected lakes, i.e. lakes that are not connected to the SWOT river network (referenced in the PLD only).

The unassigned features may be lakes that are not present in the PLD, parts of rivers that are not in the PRD, wetlands, or bright land (false detection due to topographic layover, urban areas, roads…).

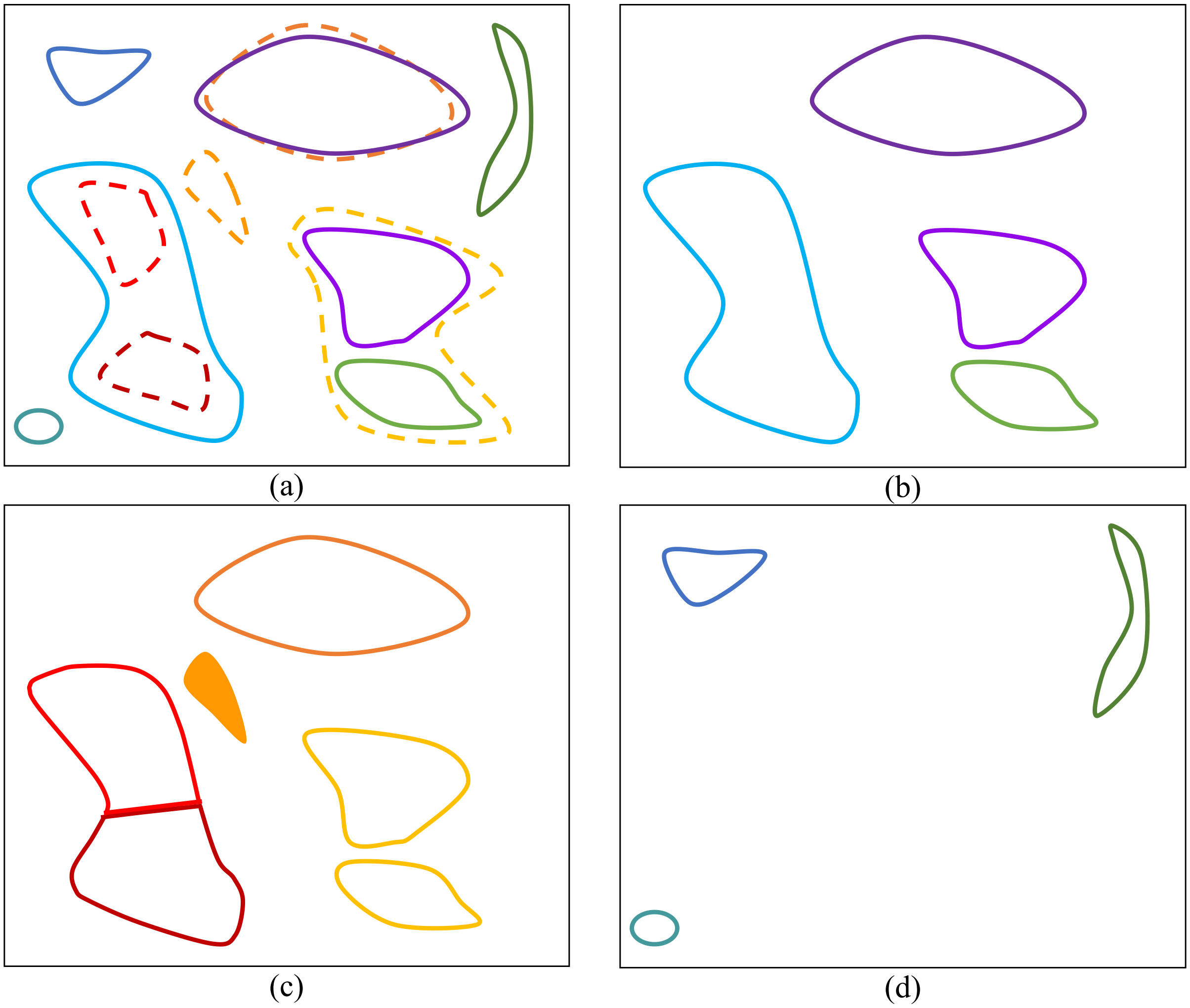

Figure 1. Illustration of how the L2_HR_LakeSP product is organized in three shapefiles. (a) All observed features (solid polygons) and PLD lakes (dashed polygons) in an area. Different colors indicate different observation identifiers or PLD identifiers. (b) Polygons of the observation-oriented lake shapefile. (c) Polygons of the PLD-oriented lake shapefile (where the unobserved PLD lake is an empty shape). (d) Polygons of the observation-oriented unassigned features shapefile.

Each shapefile consists of a set of five files with filename extensions as defined in [4]. A description of these files is provided in Table 1, below.

Table 1. Description of the files representing the L2_HR_LakeSP shapefiles.

File	Name	Description
1	Main shapefile (.shp)	Provides coordinates (polygons shape) delineating boundaries of observed water bodies and the boundaries of any island within them
2	Index file (.shx)	Stores the index of each polygon in the .shp file
3	Attributes file (.dbf)	Provides attributes for each polygon in the .shp file
4	Projection file (.prj)	Provides map projection and coordinate reference description
5	Metadata file (.shp.xml)	Provides metadata for the product

Each file in the shapefile set has the same filename prefix. The .shp file contains the basic geometry of the detected water body, computed from SWOT observations. The .dbf file contains the SWOT observations of water surface elevation (WSE or height), area, and other attributes along with information from the PLD as described in Section 4 (except for the shapefile for unassigned features which does not contain PLD information). The .prj file contains a projection description, using a well-known text (WKT) representation of coordinate reference systems (CRS). The .shp.xml file, which is not defined by the Esri specification [4], carries metadata applicable across lake shapefiles (e.g., SWOT pass number), and per-attribute metadata (e.g., units for each attribute). The format of the .shp.xml file is described in Appendix B.

Note that the use of the term "attributes" in this document follows the shapefile nomenclature in referring to the variables associated with each feature in the .shp file. The term should not be confused with attributes as typically used in the context of NetCDF files. This document uses the term "attributes" in reference to the contents of the .dbf file and uses the term "metadata" in reference to characteristics of each attribute of the entire shapefile. Therefore, as an example, in the context of this document, the SWOT-observed WSE would be an attribute of a given lake, and the metadata of the WSE attribute would indicate that the value is given in meters as the unit of measure.

Note that the names of attributes in shapefiles can be no more than 10 characters, which explains the abbreviated or truncated names of many lake attributes. Owing to this restriction, the naming conventions of attributes in the L2_HR_LakeSP product sometimes differ from the naming conventions of similar variables in other NetCDF-based SWOT data products.

3.3 File Naming Convention

The L2_HR_LakeSP product adopts the following file naming convention:

SWOT_L2_HR_LakeSP_<FileIdentifier>_<CycleID>_<PassID>_<ContinentID>_<RangeBeginningDateTime>_<RangeEndingDateTime>_<CRID>_<ProductCounter>.<extension>

The <*FileIdentifier*> above indicates whether the file is part of the observation-oriented lake shapefile (Obs), the PLD-oriented lake shapefile (Prior), or the shapefile for unassigned features (Unassigned).

The two-letter <ContinentID> above is described in Table 3.

The <CRID> above contains the composite release identifier. It contains the version code of the data product, which changes if the processing software and/or the PLD is updated.

The <extension> above indicates which of the five parts of the shapefile it is (.shp, .shx, .dbf, .prj, .shp.xml), per Section 3.2.

Example:

SWOT_L2_HR_LakeSP_Obs_001_037_NA_20210612T072103_20210612T075103_PGA2_03.shp

SWOT_L2_HR_LakeSP_Prior_001_037_NA_20210612T072103_20210612T075103_PGA2_03.shp

SWOT_L2_HR_LakeSP_Unassigned_001_037_NA_20210612T072103_20210612T075103_PGA2_03.shp

3.4 Spatial Sampling and Resolution

For simplicity, the term "lake" refers to both lakes identified in the PLD and unassigned features in the remainder of Section 3, except when they are specifically distinguished.

The content of each shapefile is a collection of records of shape type *polygon*. The polygon corresponds to the concave hull of the pixels after height-constrained geolocation [5] (Figure 2). The polygons are obtained by vectorizing PIXC (PIXCVec) edge pixels and can therefore be considered to have similar posting (in the order of 20 m in average) and resolution [6].

Lakes may be partially observed if they are located at the near or far range edges of the swath (Figure 3). The observation can also be incomplete because the water radar response is too weak to be detected as water for all or part of the lake (so-called "dark water" conditions). The *partial_f* and *dark_frac* flags inform the user of the quality degradation due to these two factors.

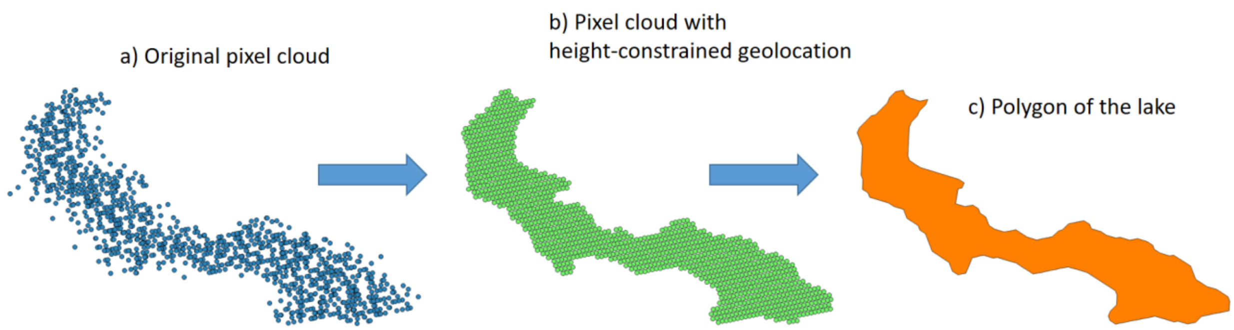

Figure 2. Synthetic scene showing that the shapefile geometry of the lake product is a polygon (c), whose shape is computed from the PIXCVec pixel cloud with height-constrained geolocation (b), rather than the more noisy PIXC pixel cloud (a).

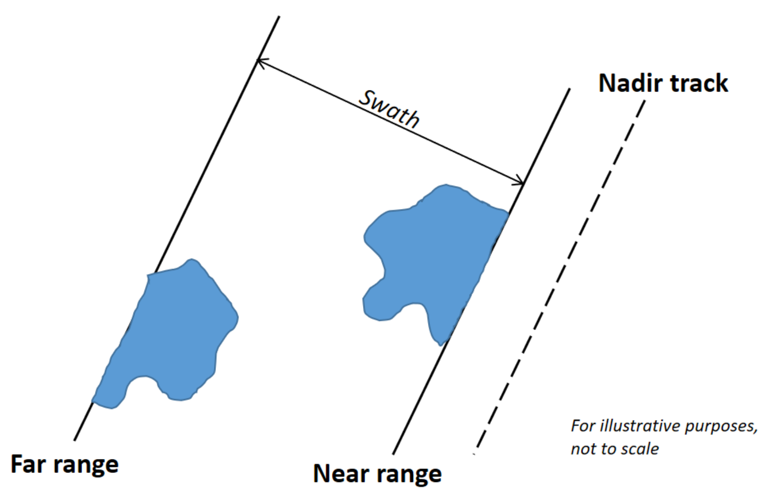

Figure 3. Example of partially observed lakes (in blue), due to their location at the edges of the swath (near and far range).

3.5 Temporal Organization

Each lake record is associated to a single time-tag corresponding to the average time-tag of all measurements contributing to the record.

The records are not strictly time-ordered in any of the three shapefiles.

3.6 Spatial Organization

Each record corresponds to a lake referenced by a geolocated polygon. This polygon is composed of one or more “rings” (using ESRI terminology); one outer ring defining the lake’s outer edge, and possibly also inner rings, delineating the contours of islands within the lake. The lake surface corresponds to the area inside the outer ring and outside any inner rings (Figure 4).

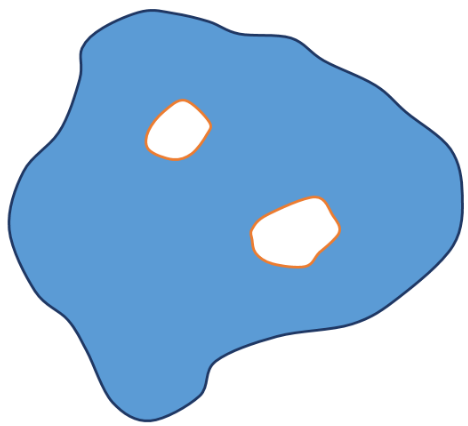

Figure 4. The lake surface (in blue) corresponds to the area comprised between the outer ring (i.e. the lake boundary, in dark blue) and any inner rings (i.e. possible islands, in orange).

In the observation-oriented L2_HR_LakeSP lake shapefile and unassigned features shapefile, the records are written in the order in which they are processed, and are not geographically ordered.

The records of the PLD-oriented L2_HR_LakeSP lake shapefile are ordered by increasing lake identifier (section 4.2).

3.7 Volume

Table 2 provides the expected volume of the individual shapefiles composing the L2_HR_LakeSP product.

The values provided in Table 2 are based on the following assumptions:

- The .dbf file for attributes represents **~232 bytes / lake** for LakeTile_Obs and LakeTile_Prior, and **~149 bytes / feature** for LakeTile_Unassigned
- The size of the .shp shape file is **144 + 4*[number of lakes] + 16*[number of lakes]*[number of points per lake] bytes**
- The .shx index file represents **100 + 8*[number of lakes] bytes**

- The number of points per polygon is considered to be ~100 on the average. As an example, in the lake *a priori* database over Europe, there are ~487000 lake polygons, with a median number of 23 points, a mean of 51 points, and a maximum of 82138 points; therefore, **100** points seems to be a conservative average estimate (i.e. with some margin).
- For the computation of the volume per granule (i.e. pass/continent), the following conservative numbers of lakes over a pass/continent granule were used:
    - **~30000 lakes** as a mean
    - **~350000 lakes** as a maximum
- As a first guess, the number of unassigned features is assumed to be the same as the number of observed PLD lakes.

Table 2. Description of the data volume of each file of L2_HR_LakeSP product.

Shapefile	Name	Expected Mean Volume/Granule (MB/pass/continent)	Maximum Volume/Granule (MB/pass/continent)
1	Obs shapefile (all files in Table 1)	53	616
2	Prior shapefile (all files in Table 1)	53	616
3	Unassigned shapefile (all files in Table 1)	50	587
	Total	156	1819

4 Qualitative Description

The L2_HR_LakeSP vector product is derived from the KaRIn measured height, geolocation, and classification data in the L2_HR_PIXC product [6]. In the description to follow, the L2_HR_PIXC products that correspond to the area of the L2_HR_LakeSP product granule are referred to as the pixel cloud, and individual pixel cloud array elements as pixels. The classification information from the pixel cloud distinguishes water pixels from land pixels (and between different types of water and land) [6]. As discussed in the Level 2 KaRIn High Rate Lake Single Pass Algorithm Theoretical Basis Document (ATBD) [5], this information is used to aggregate high-resolution pixel cloud data to the L2_HR_LakeSP vector products. That is, the pixels not already associated to a known river feature from the PRD are aggregated into separate entities. An exception is for lakes connected to a river topology, which are contained in both the L2_HR_RiverSP and the L2_HR_LakeSP products. Once the pixels are assigned, ensemble measurement quantities for each feature are computed from the pixels that were assigned to the feature. Storage change is only computed for features identified as lakes in the PLD, and is represented both in the observation-oriented lake shapefile and in the PLD-oriented lake shapefile. Unassigned features are stored in a separate observation-oriented shapefile.

The files that make up the shapefile format are described in Section 3.2. The format of the .shp file is specified in [4]. The .shp file provides geolocated polygons (latitudes and longitudes) defining lake boundaries as well as any island in it, derived from SWOT measurements. There is one record for each observed lake in the granule in the observation-oriented lake shapefile, one record for each PLD-lake covered by the granule in the PLD-oriented lake shapefile, and one record for each remaing observed feature in the shapefile for unassigned features. Measured or observed (the terms are used interchangeably in this document) values for lake attributes are mostly calculated from pixel attributes as “representative values”. The methods for calculating each attribute are given in [5]. Each record in the .dbf file contains attributes that can be conceptually grouped in the subsections below.

The following conventions are applied to the attribute names:

- Prefix “p_” indicates that information is taken from the PLD,
- Suffix “_c” indicates a correction,
- Suffix “_f” indicates a flag,
- Suffix “_u” indicates an uncertainty. Unless otherwise stated, all uncertainties represent one-sigma or 68th-percentile uncertainty estimates.

Attributes are tagged explicitly as “Basic” or “Expert” in the product metadata. This tag allows the distribution agent to make a subset file of Basic items for users who need only that information. Basic items are intended for users who will use the information derived from the KaRIn measurements as provided. Expert items are intended for users who are interested in the details of how the KaRIn measurements were derived and who may use detailed information for their own customized processing. Basic items include time, the main hydrological attributes (with uncertainties), flags, and related items from the PLD. Expert items include additional measurements, instrument and correction information. Details of the attributes are given in Section 5.

Unless otherwise specified, quantities are given in SI (MKS) units. Note that surfaces are given in km^2 rather than m^2, and likewise volumes are given in km^3 rather than m^3.

4.1 Observation-oriented lake file

4.1.1 Identifiers

The SWOT L2_HR_LakeSP product stores all the features detected as water and not processed as regular river reaches. These features may be lakes referenced in the PLD, including reservoirs that are also referenced in the PRD, or unassigned features not found neither in the PLD nor in the PRD. Moreover, note that the data used to construct the PLD (such as high-resolution optical satellite imagery) may have observed these lakes during a specific season. Therefore, it is possible that a SWOT-observed lake may correspond to a drier or more flooded state of a lake than represented in the PLD. Figure 3 shows one such example, where two lakes defined in the PLD (orange shapes) are encompassed by the actual SWOT-observed water body shape (blue shape).

To accommodate this situation, two identifier attributes are included, *obs_id* and *lake_id*. The *obs_id* attribute identifies a detected water feature and the associated L2_HR_PIXC pixel cloud tile [6]. The *lake_id* attribute identifies one or more lakes in the PLD intersecting the observed lake. It provides the link between these SWOT-observed lake locations and their corresponding entries in the PLD. Note that the *lake_id* attribute is analogous to the *reach_id* and *node_id* attributes in the river vector (L2_HR_RiverSP) product [2]. For the observation-oriented lake file the identifiers and associated attributes are as follows:

- *obs_id* (Basic): Identifier of the observed water body. It is unique to each detected water feature and observation within the cycle and pass. The format of the identifier is a 13-character string of the form **CBBTTTSNNNNNN** where **C**=continent code, **BB**=basin code, **TTT**= L2_HR_PIXC tile number within the pass, **S**=swath side (R for Right and L for Left), and **NNNNNN**=lake counter in the L2_HR_PIXC tile. For a lake spanning multiple tiles, the *obs_id* identifier corresponds to the L2_HR_PIXC tile that provided the majority of the pixels for the observed lake.
- *lake_id* (Basic): Identifier of lake(s) in the PLD which intersect the observed lake. The format of the identifier is a 10-character string of the form **CBBNNNNNNT** where **C**=continent code, **BB**=basin code, **NNNNNN**=lake counter in the basin, and **T**=water body type. If the observed lake intersects more than one prior lake, this attribute consists of a list of all the *lake_id* identifiers, separated by a semicolon character; the identifiers are ordered by decreasing overlapping area, i.e. the first PLD identifier in the list corresponds to the PLD lake having the largest overlapping area with the observed lake (cf. Figure 5).
- The attribute *overlap* provides, for each PLD lake listed in *lake_id*, the fraction (integer percentage) of the observed lake that is overlapped by the PLD lake, i.e. the overlap area divided by the total area of the observed lake.

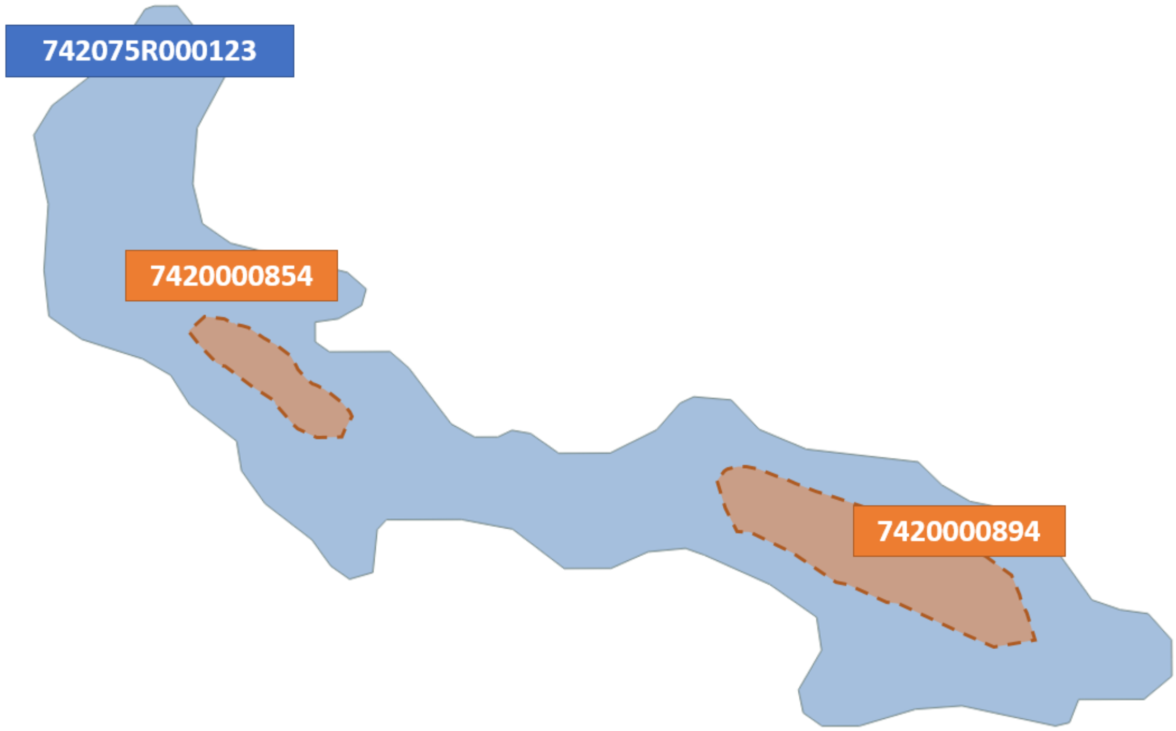

Figure 5. Observed lake (in blue) and associated PLD lakes (in orange). Observed lakes are considered associated to PLD lakes if their polygons intersect. In this case, *obs_id* = "742075R000123" meaning the continent code is 7=North America, the basin code is 42, the tile number is 75 right swath (R), and the number of the lake within the tile is 123. For the *lake_id*, note that there are two associated lakes (type=4) in the PLD, indexed as lake 85 and lake 89 in basin 42. The overlapping area between PLD lake "7420000894" and observed lake is larger than the overlapping area between PLD lake "7420000854" and observed lake. Therefore the associated *lake_id* = "7420000894; 7420000854".

Both of these identifiers are based on the Pfafstetter coding system [7] that is based on the topology of the river network. The code allows digits 1-9 at each hierarchy level. Continent code (**C**) and water body type (**T**) codes are provided in Table 3 and Table 4, respectively.

As indicated in Table 4, lakes are separated into two types, connected lakes (T=3) and disconnected lakes (T=4). Note that lake water bodies that are connected to the river topology (T=3) are included both in the L2_HR_LakeSP product and in the L2_HR_RiverSP product.

Table 3. Continent codes for the *obs_id* and *lake_id* attributes, and continent IDs for the filename.

Continent Code (C)	Continent	Continent ID ID
1	Africa	AF
2	Europe and Middle East	EU
3	Siberia	SI
4	Central and South-East Asia	AS
5	Australia and Oceania	AU
6	South America	SA
7	North America and Caribbean	NA
8	North American Arctic	AR
9	Greenland	GR

Table 4. Water body type codes for the *obs_id* and *lake_id* attributes.

Type Code (T)	Water Body Type
1	River *(not used in this product)*
2	Disconnected lake
3	Connected lake
4	Dam *(not used in this product)*
5	No topology *(not used in this product)*

The continent code (**C**) is Level 1 in the Pfafstetter code. As indicated in the template above, *obs_id* and *lake_id* values are based upon Pfafstetter Level 3, leading to 3 digits (**CBB**). Note that the continent codes in Table 3 are consistent with the continent coding used in the HydroBASINS product [8].

Figure 6 shows an example of what the coding might look like at Level 1 ($C = 7$) for the basin encompassing Mississippi (first level $B = 4$). Within each basin level, the lake is numbered with 000001 to a maximum of 999999 (i.e., a zero-padded six-digit number (TBD), represented as **NNNNNN**). Note that in the *obs_id* attribute name, this six-digit number refers to the L2_HR_PIXC tile number within the continent pass.

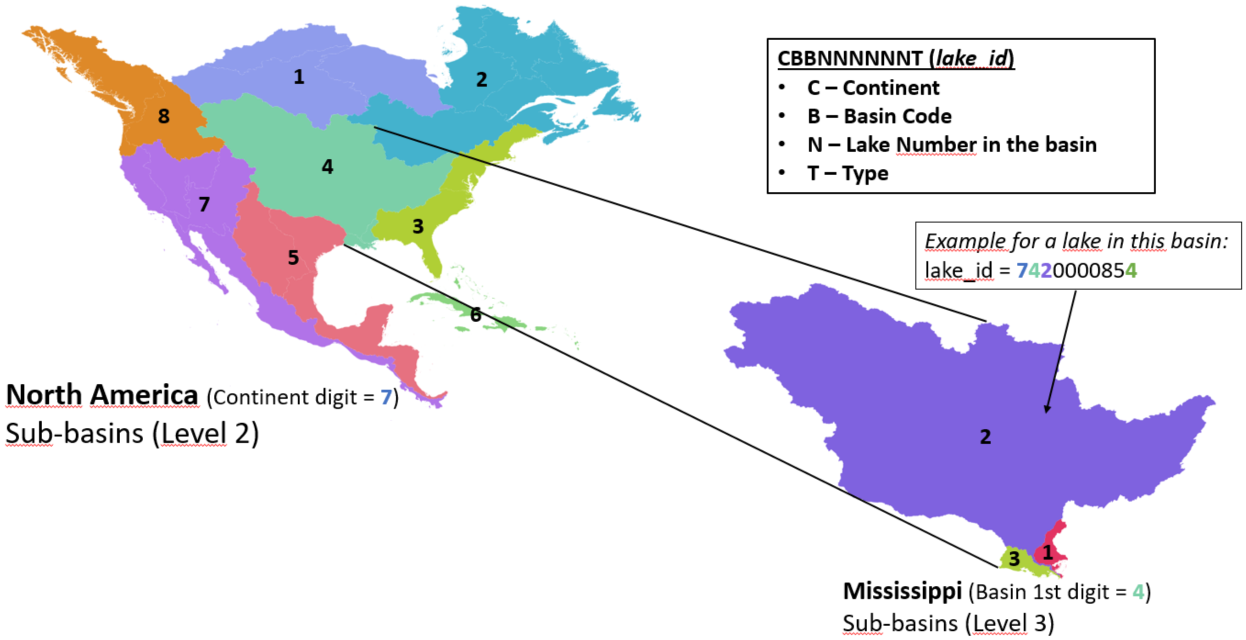

Figure 6. Example of the Pfafstetter basin coding system over North America.

4.1.2 Time

Time tags for each polygon are provided in UTC and TAI time scales using the attributes *time* and *time_tai*, respectively:

- *time*: Time in UTC time scale (seconds since January 1, 2000 00:00:00 UTC which is equivalent to January 1, 2000 00:00:32 TAI),
- *time_tai*: Time in TAI time scale (seconds since January 1, 2000 00:00:00 TAI, which is equivalent to December 31, 1999 23:59:28 UTC).

The attribute *time* has a metadata field named *tai_utc_difference*, which represents the difference between TAI and UTC (i.e., total number of leap seconds) at the time of the first measurement record in the products granule.

- *time_tai[0] = time[0] + tai_utc_difference*

The above relationship holds true for all measurement records unless an additional leap second occurs within the time span of the products granule. To account for this, the variable *time* also has a metadata field named *leap_second* which provides the date at which a leap second might have occurred within the time span of the products granule. The variable *time* exhibits a jump when a leap second occurs. If no additional leap second occurs within the time span of the products granule *time: leap_second* is set to the UTC time at which the leap second occurs.

Table 5 below provides some examples for the values of *time*, *time_tai*, and *tai_utc_difference*. With this approach, the value of *time* has a 1 second regression during a leap second transition, while *time_tai* is continuous. That is, when a positive leap second is inserted, two different instances have the same value for the variable *time*, making time non-unique by itself; the difference between *time* and *time_tai*, or the *tai_utc_difference* and *leap_second* fields, can be used to resolve this.

Table 5. Examples of UTC and TAI dates, and corresponding *time*, *time_tai*, and *tai_utc_difference.*

UTC Date	TAI Date	time	time_tai	tai_utc_difference
January 1, 2000 00:00:00	January 1, 2000 00:00:32	0.0	32.0	32
December 31, 2016 23:59:59	January 1, 2017 00:00:35	536543999.0	536544035.0	36
December 31, 2016 23:59:59.5	January 1, 2017 00:00:35.5	536543999.5	536544035.5	36
December 31, 2016 23:59:60	January 1, 2017 00:00:36	536543999.0	536544036.0	37
January 1, 2017 00:00:00	January 1, 2017 00:00:37	536544000.0	536544037.0	37
January 1, 2017 12:00:00	January 1, 2017 12:00:37	536587200.0	536587237.0	37

The UTC time corresponding to the numeric *time* attribute is also given as a string attribute (*time_str*): YYYY-MM-DDThh:mm:ssZ ('Z' suffix to indicate UTC time).

4.1.3 Measured hydrological parameters

The basic hydrological attributes of the SWOT measurement include WSE (*wse*), water surface area (*area_total*), and storage change (*delta_s_l* and *delta_s_q*).

Two additional (expert) area attributes (*area_detct* and *area_wse*) are provided for information and for quality-assessment purposes. *area_detct* includes the actual SWOT-detected water surface area, including open water and area near land-water boundaries [5]. The basic attribute *area_total* is the sum of *area_detct* and the area of any "dark water" (the area of water that was not observed directly by SWOT owing to a low radar echo level, which can occur over very smooth water surfaces, or by significant attenuation of the radar signal due to propagation through rain). Areas of dark water areidentified in ground processing through the use of a prior water probability map [6].

The attribute *area_wse* represents the water surface area over which the SWOT measurements of height contribute to the reported WSE (*wse*) for the lake. The value of *area_wse* may be less than *area_total* if some of the measurements fail validity checks during processing.

Each basic quantity has an associated uncertainty. The methods for calculating the lake quantities and associated uncertainties from the pixel values are given in the ATBD [5]. For WSE, the random-only component of the total uncertainty (*wse_r_u*) is provided in addition to the total uncertainty (*wse_u*). The random-only component here is taken to be the component that is independent between lakes, not including systematic errors that would be common from lake to lake within a granule. The WSE systematic uncertainty component can be computed from sqrt($(wse\_u)^2 - (wse\_r\_u)^2$).

The WSE given in the product is reported with respect to the provided model of the geoid (*geoid_hght*), and after using models to accounts for the effects of tides (see Section 4.1.6). Specifically, if H represents the geocentric height of the water surface with respect to the reference ellipsoid after applying corrections for media delays (see Section 4.1.7) and the crossover calibration (see Section 4.1.8), then *wse* is computed as follows:

$$wse = H - geoid\_hght - solid\_tide - load\_tidef - pole\_tide$$

Attributes are tagged as basic or expert, as indicated in the table in Section 5 (tag_basic_expert).

- *wse* (Basic): Average water surface elevation of the lake, relative to the provided model of the geoid (*geoid_hght* attribute, see Section 4.1.6), with corrections for media delays (wet and dry troposphere, and ionosphere) and tidal effects (*solid_tide*, *load_tidef*, and *pole_tide*) applied.
- *wse_u* (Basic): Total uncertainty (random and systematic) in the lake WSE. The value includes uncertainties of corrections and references.
- *wse_r_u* (Expert): Random-only component of the uncertainty in the lake WSE.
- *wse_std* (Basic): Standard deviation of the water surface elevation of all the pixels composing the lake. Note that this value is computed after removing the contributions of the geoid and tide terms (see the *wse* variable), whereas the heights of pixels in the L2_HR_PIXC product are given with respect to the ellipsoid and without tide corrections. This parameter is computed only for lakes > 5000ha (TBC).
- *area_total* (Basic): Total estimated water surface area, including *area_detct* and any dark water that was not detected as water in the SWOT observation, but identified through the use of a prior water likelihood map.
- *area_tot_u* (Basic): Total uncertainty (random and systematic) in the total estimated water surface area *area_total*.
- *area_detct* (Expert): Actual SWOT-detected water surface area, including open water and water near land.
- *area_det_u* (Expert): Total uncertainty (random and systematic) in the surface area of the detected water pixels.
- *layovr_val* (Expert): TBD value indicating an estimate of the WSE error due to layover.
- *xtrk_dist* (Basic): Distance of the lake polygon centroid from the spacecraft nadir track; this value is computed using a local spherical Earth approximation. A negative value indicates that the lake is on the left side of the swath, relative to the spacecraft velocity vector. A positive value indicates that the lake is on the right side of the swath.

4.1.4 Storage change

Storage change is computed only for observed lakes linked to one or more PLD lake. It is derived from SWOT measurements with respect to the reference area and height of the PLD lake.

Note that if there is only one PLD lake associated to the observed lake, these PLD attributes are provided in *p_area* and *p_height* attributes, respectively (Section 4.1.9). If there is more than one PLD lake, the PLD attributes provided in the product and listed in Section 4.1.9 are those of the PLD lake having the largest overlap with the observed lake. Therefore, the user should refer directly to the PLD to retrieve all the reference areas and heights from which the storage change is computed.

The storage change is computed with two different methods [5]:

- *delta_s_l*, *ds_u_l* (Basic): Storage change and associated uncertainty, computed by the linear method.
- *delta_s_q*, *ds_u_q* (Basic): Storage change and associated uncertainty, computed by the quadratic method.

Note that storage change is first computed with respect to the *p_max_wse* and *p_max_area* of the PLD lake, and then corrected by *p_ref_ds* to respresent the storage change relative to *p_ref_date*, i.e., the date of the first valid measurement. For the date *p_ref_date*, the storage change attributes *delta_s_l* and *delta_s_q* will therefore be zero.

4.1.5 Quality indicators

Flags indicating conditions that affect data quality are given as basic attributes. In general, flag values of zero indicate good data.

- *quality_f* (Basic): Summary quality indicator for the lake measurement. Values of 0 and 1 indicate nominal and off-nominal measurements, respectively.
- *dark_frac* (Expert): Fraction of the lake total area (*area_total*) covered by dark water, equal to 1-(area_detct/area_total). This value is typically between 0 and 1, with 0 indicating no dark water and 1 indicating 100% dark water. However, the value may be outside the range from 0 to 1 due to noise in the underlying area estimates.
- *ice_clim_f* (Basic): Climatological ice cover flag indicating whether the lake is ice-covered on the day of the observation based on climatological ice coverage [9] (not the SWOT measurement). Values of 0, 1, and 2 indicate that the lake is likely not ice covered, may or may not be partially or fully ice covered, and likely fully ice covered, respectively. In case of an observed lake associated to more than one PLD lake, the resulting flag is a combination of the flags of all the prior lakes. For example, if the observed lake intersects two PLD lakes, one being fully ice-covered and the other not, the resulting flag is partially ice-covered.
- *ice_dyn_f* (Basic): Dynamic ice cover flag indicating whether the lake is ice-covered on the day of the observation based on analysis of external optical satellite data [9] (not the SWOT measurement). Values of 0, 1, and 2 indicate that the lake is not ice covered, partially ice covered, and fully ice covered, respectively. Due to the latency of computing the dynamic ice flag, this value may be completely null filled in some processing versions of the data product. When available, *ice_dyn_f* is likely to be more reliable than *ice_clim_f* given that it is based on optical satellite observations representative of the surface conditions at the time of the SWOT observation.
- *partial_f* (Basic): Flag that indicates partial lake coverage. The flag is 0 if the observed lake is fully covered by a half-swath; the flag is 1 if the observed lake hits the right or left edge of the half-swath, and therefore, that a part of it may be lost.
- *xovr_cal_q* (Basic): Flag that indicates the quality of the cross-over calibration. TBD.

4.1.6 Geophysical references

The geoid height, from a model, in meters above the reference ellipsoid (defined by the .prj file) is a basic attribute. This information enables the user to convert the observed WSE to a different representation.

Expert attributes provide the tide heights, from models, that were used to calculate the *wse* attribute. Note that while the model solution used to account for the effect of the ocean tide loading on the Earth's crust is provided in the attribute *load_tidef*, a second model solution (*load_tideg*) is provided for users who desire to swap these models. Each geophysical reference value provided in the product is computed as the average over all pixels in the associated feature [5].

The associated geophysical reference parameters include:

- *geoid_hght* (Basic): Model for geoid height above the reference ellipsoid whose parameters are given in the.prj file. The geoid model is EGM2008 [10]. The geoid model includes a correction to refer the value to the mean tide system (i.e., it includes the zero-frequency permanent tide).
- *solid_tide* (Expert): Model for the solid Earth (body) tide height. The reported value is calculated using the Cartwright/Taylor/Edden [11] [12] tide-generating potential coefficients and consists of the second and third degree constituents. The permanent tide (zero frequency) is not included.
- *pole_tide* (Expert): Model for the surface height displacement from the geocentric pole tide. The value is the sum total of the contribution from the solid-Earth (body) pole tide height *[13]* and a model for the load pole tide height [14]. The value is computed using the reported Earth pole location after correction for a linear drift *[15]*: in milliarcsec,
$$Xp = 55.0 + 1.677dt$$
$$Yp = 320.5 + 3.46dt$$
where dt is years since 2000.0.
- *load_tidef* (Expert): Model for geocentric surface height displacement from the load tide. The value is from the FES2014b ocean tide model [16]. The value is used to compute *wse*.
- *load_tideg* (Expert): Model for geocentric surface height displacement from the load tide. The value is from the GOT4.10c ocean tide model [17]. To compute *wse* with this model, add *load_tidef* to *wse* and subtract *load_tideg*.

4.1.7 Geophysical range corrections

Model-based corrections for the wet and dry troposphere and the ionosphere contributions to the measured range are provided for each lake as expert attributes. The reported values are averages over the values of the pixels in the associated feature [5]. Additional details on these media delays are provided in [6]. Note that while these media delays are corrected during processing along the slanted (non-vertical) radar signal propagation paths, they are provided in these attributes as equivalent vertical quantities after applying a cross-track-dependent obliquity factor. The additional path delay relative to free space results is a negative correction value that is added to the uncorrected range. However, a decrease in the measured range gives an increase in the measured height. Consequently, adding the reported correction terms to the reported *wse* value results in the uncorrected reach WSE. Model-based corrections are based on SWOT-independent information from the European Centre for Medium-Range Weather Forecasts (ECMWF) and Jet Propulsion Laboratory (JPL) Global Ionosphere Maps (GIM). The sources of the model data used for these corrections are given in the metadata provided in the .shp.xml file (see Section 5.1.3).

- *dry_trop_c* (Expert): Model-based equivalent vertical dry tropospheric path delay correction. This value is computed using surface pressure from the ECMWF numerical weather model.
- *wet_trop_c* (Expert): Model-based equivalent vertical wet tropospheric path delay correction. This value is computed from the ECMWF numerical weather model.
- *iono_c* (Expert): Equivalent vertical ionospheric path delay correction from the JPL Global Ionosphere Maps (GIM) for the KaRIn Ka-band signal.

4.1.8 Instrument corrections

Instrument corrections applied to the KaRIn data are provided as expert attributes. The crossover correction is based on SWOT observations over the ocean to correct for attitude effects. Further details are supplied in the ATBD [5]. These corrections are provided so that a different or updated calibration can be applied to the lake height without regenerating the pixel cloud or lake vector products.

- *xovr_cal_c* (Expert): Height correction to *wse* computed from a combination of sea surface height crossovers between KaRIn/KaRIn measurements and KaRIn/nadir altimeter measurements on different passes within a temporal window surrounding the height measurement. This correction provides an estimate of residual errors that have not been removed with use of ancillary attitude and calibration data during processing. The correction is applied before geolocation, but it is reported in the product as an equivalent height correction. The correction term should be subtracted from the reported *wse* to obtain the uncorrected WSE.

4.1.9 Prior Lake Database (PLD) information

Information from the PLD is provided in the product with the measurements to allow easier connection or comparison. The sources, methods of generation, and accuracy are described in [3].

If an observed lake is associated to more than one PLD lake, the attributes are populated with the values corresponding to the PLD lake having the largest overlapping area with the observed lake.

All attributes are basic:

- *p_name* (Basic): Name(s) of the lake, retrieved from Open Street Map, IGN Carthage, GLWD and vMap0 databases.

- *p_grand_id* (Basic): Reservoir identifier from the Global Reservoir and Dam (GRanD) database *[23]*. Note that if the PLD lake corresponds to a reservoir, its corresponding *lake_id* identifier ends with digit 3.
- *p_max_wse* (Basic): Maximum water surface elevation (except flooding events); if only one PLD lake intersects the observed lake, this is the reference height used to compute the storage change (Section 4.1.4).
- *p_max_area (Basic):* Maximum water surface area (except flooding events); if only one PLD lake intersects the observed lake, this is the reference area used to compute the storage change (Section 4.1.4).
- *p_ref_date* (Basic): Reference date for the storage change attributes described in section 4.1.4, corresponding to the date of the first valid measurement. Note that this date does not necessarily correspond to the maximum water surface elevation and area.
- *p_ref_ds* (Basic): Reference storage change used to translate the storage change values initially computed with respect to the *p_max_wse* and *p_max_area* of the PLD lake, to the storage change relative to *p_ref_date*.
- *p_storage* (Basic): Maximum water storage value, computed between the minimum (or ground when a bathymetry is available) and maximum observed levels of the lake. This field will be filled after one year of SWOT mission.

4.2 PLD-oriented lake file

The PLD-oriented lake shapefile has the same structure and attributes as the observation-oriented lake shapefile described in section 4.1, except that it has one record per PLD lake covered by the granule (including unobserved PLD lakes), rather than one record per observed lake. There is therefore one unique *lake_id* per record, but possibly several *obs_id*s, listed by deacreasing fractional overlap as given by the attribute *overlap* (note that the computation of this attribute is different from that of the observation-oriented lake shapefile, as described below):

- *lake_id* (Basic): Principal identifier of the records of the PLD-oriented lake shapefile, corresponding to the PLD lake identifier. The format of the identifier is a 10-character string of the form **CBBNNNNNNT** where **C**=continent code, **BB**=basin code, **NNNNNN**=lake counter in the basin, and **T**=water body type.
- *obs_id* (Basic): Identifier of the observed water body. It is unique to each detected water feature and observation within the cycle and pass. The format of the identifier is a 13-character string of the form **CBBTTTSNNNNNN** where **C**=continent code, **BB**=basin code, **TTT**= L2_HR_PIXC tile number within the pass, **S**=swath side (R for Right and L for Left), and **NNNNNN**=lake counter in the L2_HR_PIXC tile. For a lake spanning multiple tiles, the *obs_id* identifier corresponds to the L2_HR_PIXC tile that provided the majority of the pixels for the observed lake. If a PLD lake intersects more than one observed lake, this attribute consists of a list of all the *obs_id* identifiers, separated by a semicolon character; the identifiers are ordered by decreasing overlapping area, i.e. the first identifier in the list corresponds to the observed lake having the largest overlapping area with the PLD lake.
- The attribute *overlap* provides, for each observed lake listed in *obs_id*, the fraction (integer percentage) of the PLD lake that is overlapped by the observed lake, i.e. the

overlap area divided by the total area of the PLD lake. Note that the computation of this attribute is different from that of the observation-oriented lake shapefile.

All other attributes are the same as in the observation-oriented lake shapefile, except that measurements are aggregated over PLD lakes rather than observed lakes. The storage change values are however the same, as they were already given per PLD lake in the in the observation-oriented lake shapefile (there repeated for each intersecting observed lake, whereas they are here given only once as each record corresponds to a PLD lake).

As illustrated in Figure 1 (c), the polygon of a record in the PLD-oriented lake shapefile is composed of one or several polygons, each of them corresponding either to an entire observed lake (for observed lakes assigned to one single PLD lake) or to part of an observed lake (for observed lakes that have been assigned to several PLD lakes). In the latter case, the split of the polygon of the observed lake is done based on the distance to the respective PLD polygons.

As opposed the observation-oriented lake shapefile, the PLD-oriented lake shapefile includes records for unobserved PLD lakes covered by the granule. The attributes of these records are set to the fill values, except for the attributes containing PLD information, and there is no polygon (empty shape).

4.3 Observation-oriented unassigned features file

The observation-oriented shapefile for unassigned features has the same structure and attributes as the observation-oriented lake shapefile described in section 4.1, except that:

- It does not have a *lake_id* attribute, nor any *overlap* attribute (Section 4.1.1).
- There are no storage change attributes (Section 4.1.4).
- It does not contain the attributes with PLD information described in Section 4.1.9.

All other attributes are computed as if the features were lakes, although they could also correspond to river portions not represented in the PRD, wetlands, or bright land. The data in this shapefile must therefore be used with caution, and the nature of an observed feature should be verified using other information sources.

Unassigned features that are observed repeatedly will be considered for inclusion when updating the PLD (and PRD) during the mission, and may this way become regular lakes (or river reaches) in reprocessed products.

5 Detailed Content

The L2_HR_LakeSP product adopts the Esri shapefile format and conventions [4]. The shapefile format stores geospatial data as primitive geometric shapes like points, polylines, and polygons representing locations, rivers, and lakes, respectively. These shapes, together with data attributes that are linked to each shape, create the representation of the geographic data. In this section a description of the information in the .dbf file is given. This information is also stored in the .shp.xml file of the lake shapefile. The .shp.xml file provides shapefile metadata information similar to what would be provided as global and per-variable attributes in a NetCDF format file. The format of the .shp.xml file is described in Appendix B.

5.1 Shapefile information

5.1.1 Dimensions

The headers of the .shp and .shx lake files give the number of records in the shapefiles. However, the .dbf file does not have an entry for the number of records. All attributes in the .dbf file are scalars (each attribute corresponds to only a single integer, floating-point value, or text). However, some attributes are multi-valued: *lake_id* and *p_name* attributes are given as character strings in a semicolon-separated list of the lakes from the PLD that intersect the observed lake, and the different names given to the lake, respectively.

5.1.2 Attributes

The attributes of the .dbf file are assigned a name and a particular data type. Note that .dbf attributes are all stored as space-separated, formatted ASCII (ANSI) character strings rather than binary data types. Table 6 summarizes the type, field width and fill value for each data type.

Table 6. Attribute data types in shapefile products.

Data Type	Description	fill value
int4	integer (4-character storage)	-999
int9	integer (9-character storage)	-99999999
float	floating point (13-character storage)	-999999999999
text	maximum 254-character storage	“no_data”

5.1.3 Metadata

The unique, descriptive metadata for each attribute (e.g., expected minimum and maximum values; e.g., the equivalent of the NetCDF attributes *valid_min*, *valid_max*) and the global metadata (e.g., SWOT pass number) generally follow the conventions defined for other SWOT products and are given in Table 7 and Table 8, respectively. Since metadata cannot be stored inside the .dbf file, the .shp.xml file will provide the metadata fields that apply to each shapefile attribute in the .dbf file. Not all metadata fields will be used for each shapefile attribute (e.g., the metadata field *leap_second* is unique to the time attributes). A description of the .shp.xml file format is given in Appendix B.

Table 7. Metadata fields used to describe shapefile attributes.

Attribute	Description
basic_expert_tag	Tag to indicate whether the attribute is considered basic or expert.
calendar	Reference time calendar.
comment	Miscellaneous information about the attribute, or the methods to generate it.
coordinates	Coordinate variables associated with the attribute.
fill_value	The value used to represent missing or undefined data.
flag_meanings	The description of the meaning of each of the elements of flag_values.
flag_values	Values of the flag attribute. Used in conjunction with flag_meanings.
institution	Institution which generates the source data for the attribute, if applicable.
leap_second	UTC time at which a leap second occurs within the time span of the data represented in the attribute.
long_name	A descriptive name that indicates the content of the attribute.
quality_flag	Names of variable quality flag(s) that are associated with this attribute to indicate its quality.
source	Data source (model, author, or instrument).
standard_name	A standard name that indicates the attribute content.
tai_utc_difference	Difference between TAI and UTC reference time.
type	Attribute type (int4, int9, float or text)
units	Units of attribute.
valid_max	Maximum theoretical value of the attribute (not necessarily the same as maximum value of actual data)
valid_min	Minimum theoretical value of the attribute (not necessarily the same as minimum value of actual data)

Table 8. Global metadata fields of the L2_HR_LakeSP product.

Attribute	Format	Description
Conventions	string	Esri conventions as given in 'ESRI Shapefile Technical Description, an ESRI White Paper, July 1998' http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
title	string	Level 2 KaRIn high rate lake single pass vector product – [Obs|Prior|Unassigned]
institution	string	Name of producing agency.
source	string	The method of production of the original data. If it was model-generated, source should name the model and its version, as specifically as could be useful. If it is observational, source should characterize it (e.g., 'Ka-band radar interferometer')
history	string	UTC time when file generated. Format is: YYYY-MM-DD hh:mm:ss : Creation
platform	string	SWOT
references	string	Provides version number of software generating product.
reference_document	string	SWOT-TN-CDM-0673-CNES
contact	string	Contact information for producer of product. (e.g., 'ops@jpl.nasa.gov').
cycle_number	short	Cycle number of the product.
pass_number	short	Pass number of the product.
continent	string	Continent the product belongs to
time_coverage_start	string	UTC time of first measurement. Format is: YYYY-MM-DDThh:mm:ss.ssssssZ

Attribute	Format	Description
time_coverage_end	string	UTC time of last measurement. Format is: YYYY-MM-DDThh:mm:ss.ssssssZ
geospatial_lon_min	double	Westernmost longitude (deg) of granule bounding box
geospatial_lon_max	double	Easternmost longitude (deg) of granule bounding box
geospatial_lat_min	double	Southernmost latitude (deg) of granule bounding box
geospatial_lat_max	double	Northernmost latitude (deg) of granule bounding box
left_first_longitude	double	Nominal swath corner longitude for the first range line and left edge of the swath (degrees_east)
left_first_latitude	double	Nominal swath corner latitude for the first range line and left edge of the swath (degrees_north)
left_last_longitude	double	Nominal swath corner longitude for the last range line and left edge of the swath (degrees_east)
left_last_latitude	double	Nominal swath corner latitude for the last range line and left edge of the swath (degrees_north)
right_first_longitude	double	Nominal swath corner longitude for the first range line and right edge of the swath (degrees_east)
right_first_latitude	double	Nominal swath corner latitude for the first range line and right edge of the swath (degrees_north)
right_last_longitude	double	Nominal swath corner longitude for the last range line and right edge of the swath (degrees_east)
right_last_latitude	double	Nominal swath corner latitude for the last range line and right edge of the swath (degrees_north)
xref_input_l2_hr_pixc_files	string	List of water mask pixel cloud files used to generate data in product.
xref_input_l2_hr_lake_tile_files	string	List of LakeTile products used to generate data in product.
xref_static_lake_db_file	string	Name of static lake a priori database file used to generate data in product.
xref_l2_hr_lake_sp_param_file	string	Name of PGE_L2_HR_LakeSP parameter file used to generate data in product.

5.2 Observation-oriented lake file attribute description

Table 9 lists the lake .dbf shapefile attributes (bold left-most column), and their associated metadata fields from Table 7. The attributes are separated into the nine categories listed in Sections 4.1 through 4.9. Appendix B contains a description of the shp.xml format that was used to generate this table.

Table 9. Attributes of the observation-oriented lake shapefile of the L2_HR_LakeSP product.

Lake ID		
obs_id		
	type	text
	long_name	identifier of the observed lake
	tag_basic_expert	Basic
	comment	Unique lake identifier within the product. The format of the identifier is CBBTTTSNNNNNN, where C=continent code, B=basin code, TTT=tile number within the pass, S=swath side, N=lake counter within the tile.
lake_id		
	type	text
	fill_value	"no_data"

	long_name	lake ID(s) from prior database
	tag_basic_expert	Basic
	comment	List of identifiers of prior lakes that intersect the observed lake. The format of the identifier is CBBNNNNNNT, where C=continent code, B=basin code, N=lake counter within the basin, T=type. The different lake identifiers are separated by semicolons.
overlap		
	type	text
	fill_value	"no_data"
	long_name	fraction of observed lake covered by each prior lake
	tag_basic_expert	Basic
	comment	List of fractions of observed lake area covered by each prior lake identified in lake_id attribute. The different fractions are separated by semicolons and refer one-to-one to the identifiers listed in the lake_id attribute.
Time		
time		
	type	float
	fill_value	-999999999999
	long_name	time (UTC)
	standard_name	time
	calendar	gregorian
	tai_utc_difference	[value of TAI-UTC at time of first record]
	leap_second	YYYY-MM-DD hh:mm:ss
	units	seconds since 2000-01-01 00:00:00.000
	tag_basic_expert	Basic
	comment	Time of measurement in seconds in the UTC time scale since 1 Jan 2000 00:00:00 UTC. [tai_utc_difference] is the difference between TAI and UTC reference time (seconds) for the first measurement of the data set. If a leap second occurs within the data set, the attribute leap_second is set to the UTC time at which the leap second occurs.
time_tai		
	type	float
	fill_value	-999999999999
	long_name	time (TAI)
	standard_name	time
	calendar	gregorian
	units	seconds since 2000-01-01 00:00:00.000
	tag_basic_expert	Basic
	comment	Time of measurement in seconds in the TAI time scale since 1 Jan 2000 00:00:00 TAI. This time scale contains no leap seconds. The difference (in seconds) with time in UTC is given by the attribute [time:tai_utc_difference].
time_str		
	type	text
	fill_value	"no_data"
	long_name	UTC time
	standard_name	time
	calendar	gregorian
	tai_utc_difference	[value of TAI-UTC at time of first record]
	leap_second	YYYY-MM-DD hh:mm:ss
	tag_basic_expert	Basic
	comment	Time string giving UTC time. The format is YYYY-MM-DDThh:mm:ssZ, where the Z suffix indicates UTC time.
Measured Hydrology Parameters		

wse		
	type	float
	fill_value	-999999999999
	long_name	lake-averaged water surface elevation with respect to the geoid
	units	m
	valid_min	-1000
	valid_max	100000
	tag_basic_expert	Basic
	comment	Lake-averaged water surface elevation, relative to the provided model of the geoid (geoid_hght), with corrections for media delays (wet and dry troposphere, and ionosphere), crossover correction, and tidal effects (solid_tide, load_tidef, and pole_tide) applied.
wse_u		
	type	float
	fill_value	-999999999999
	long_name	total uncertainty in lake water surface elevation
	units	m
	valid_min	0
	valid_max	100
	tag_basic_expert	Basic
	comment	Total one-sigma uncertainty (random and systematic) in the lake WSE, including uncertainties of corrections and references.
wse_r_u		
	type	float
	fill_value	-999999999999
	long_name	random-only uncertainty in the height water surface elevation
	units	m
	valid_min	0
	valid_max	100
	tag_basic_expert	Expert
	comment	Random-only component in the lake water surface elevation, including uncertainties of corrections and references, and variation about the fit.
wse_std		
	type	float
	fill_value	-999999999999
	long_name	standard deviation of pixels wse
	units	m
	valid_min	-1000
	valid_max	100000
	tag_basic_expert	Basic
	comment	Standard deviation of the water surface elevation of all the pixels composing the lake. Note that this value is therefore with respect to the provided model of the geoid (*geoid_hght* attribute) whereas the height of pixels is given with respect to the ellipsoid. This parameter is computed only for large lakes (> 5000ha).
area_total		
	type	float
	fill_value	-999999999999
	long_name	total water area with estimate of dark water
	units	km^2
	valid_min	0
	valid_max	200000
	tag_basic_expert	Basic

	comment	Total estimated area, including dark water that was not detected as water in the SWOT observation but identified through the use of a prior water likelihood map.
area_tot_u		
	type	float
	fill_value	-999999999999
	long_name	uncertainty in total water area
	units	km^2
	valid_min	0
	valid_max	200000
	tag_basic_expert	Basic
	comment	Total uncertainty (random and systematic) in the total water area.
area_detct		
	type	float
	fill_value	-999999999999
	long_name	area of detected water pixels
	units	km^2
	valid_min	0
	valid_max	200000
	tag_basic_expert	Expert
	comment	Aggregation of used detected pixels area.
area_det_u		
	type	float
	fill_value	-999999999999
	long_name	uncertainty in area of detected water
	units	km^2
	valid_min	0
	valid_max	200000
	tag_basic_expert	Expert
	comment	Total uncertainty (random and systematic) in the area of detected water pixels.
layovr_val		
	type	float
	fill_value	-999999999999
	long_name	metric of layover effect
	units	TBD
	valid_min	TBD
	valid_max	TBD
	tag_basic_expert	Expert
	comment	Value indicating an estimate of the height error due to layover.
xtrk_dist		
	type	float
	fill_value	-999999999999
	long_name	distance of lake polygon centroid to the satellite ground track
	units	m
	valid_min	-75000
	valid_max	75000
	tag_basic_expert	Basic
	comment	Distance of centroid of polygon delineating lake boundary to the satellite ground track. A negative value indicates the left side of the swath, relative to the spacecraft velocity vector. A positive value indicates the right side of the swath.
Storage change		
delta_s_l		
	type	float
	fill_value	-999999999999

	long_name	storage change computed by linear method
	units	km^3
	valid_min	-1000
	valid_max	1000
	tag_basic_expert	Basic
	comment	Storage change with regards to the reference area and height from PLD; computed by the linear method.
ds_l_u		
	type	float
	fill_value	-999999999999
	long_name	uncertainty in storage change computed by linear method
	units	km^3
	valid_min	-1000
	valid_max	1000
	tag_basic_expert	Basic
	comment	Uncertainty in storage change computed by linear method.
delta_s_q		
	type	float
	fill_value	-999999999999
	long_name	storage change computed by quadratic method
	units	km^3
	valid_min	-1000
	valid_max	1000
	tag_basic_expert	Basic
	comment	Storage change with regards to the reference area and height from PLD; computed by the quadratic method.
ds_q_u		
	type	float
	fill_value	-999999999999
	long_name	uncertainty in storage change computed by quadratic method
	units	km^3
	valid_min	-1000
	valid_max	1000
	tag_basic_expert	Basic
	comment	Uncertainty in storage change computed by quadratic method.
Quality Indicators		
quality_f		
	type	int4
	fill_value	-999
	long_name	summary quality indicator for lake measurement
	flag_meanings	good bad
	flag_values	0 1
	valid_min	0
	valid_max	1
	tag_basic_expert	Basic
	comment	Summary quality flag for the lake measurement. Values of 0 and 1 indicate nominal and off-nominal measurements.
dark_frac		
	type	float
	fill_value	-999999999999
	long_name	fractional area of dark water
	units	1
	valid_min	0

	valid_max	1
	tag_basic_expert	Expert
	comment	Fraction of lake area_total covered by dark water. The value is between 0 and 1.
ice_clim_f		
	type	int4
	fill_value	-999
	long_name	climatological ice cover flag
	standard_name	status_flag
	source	Yang et al. (2020)
	flag_meanings	no_ice_cover uncertain_ice_cover full_ice_cover
	flag_values	0 1 2
	valid_min	0
	valid_max	2
	tag_basic_expert	Basic
	comment	Climatological ice cover flag indicating whether the lake is ice-covered on the day of the observation based on external climatological information (not the SWOT measurement). Values of 0, 1, and 2 indicate that the lake is likely not ice covered, may or may not be partially or fully ice covered, and likely fully ice covered, respectively.
ice_dyn_f		
	type	int4
	fill_value	-999
	long_name	dynamical ice cover flag
	standard_name	status_flag
	source	Yang et al. (2020)
	flag_meanings	no_ice_cover partial_ice_cover full_ice_cover
	flag_values	0 1 2
	valid_min	0
	valid_max	2
	tag_basic_expert	Basic
	comment	Dynamic ice cover flag indicating whether the lake is ice-covered on the day of the observation based on analysis of external optical satellite data. Values of 0, 1, and 2 indicate that the lake is not ice covered, partially ice covered, and fully ice covered, respectively.
partial_f		
	type	int4
	fill_value	-999
	long_name	partially covered lake flag
	flag_meanings	covered partially_covered
	flag_values	0 1
	valid_min	0
	valid_max	1
	tag_basic_expert	Basic
	comment	Flag that indicates only partial lake coverage. 0= Indicates that the observed lake is entirely covered by the swath. 1= Indicates that the observed lake is partially covered by the swath.
xovr_cal_q		
	type	int4
	fill_value	-999
	long_name	quality of the cross-over calibrations
	flag_meanings	TBD
	flag_masks	TBD
	flag_values	TBD

	valid_min	0
	valid_max	TBD
	tag_basic_expert	Basic
	comment	Quality of the cross-over calibration.
Geophysical References		
geoid_hght		
	type	float
	fill_value	-999999999999
	long_name	geoid height
	standard_name	geoid_height_above_reference_ellipsoid
	source	EGM2008
	institution	GSFC
	units	m
	valid_min	-150
	valid_max	150
	tag_basic_expert	Basic
	comment	Lake-averaged geoid model height above the reference ellipsoid. The value is computed from the EGM2008 geoid model with a correction to refer the value to the mean tide system (i.e., includes the zero-frequency permanent tide).
solid_tide		
	type	float
	fill_value	-999999999999
	long_name	solid Earth tide height
	source	Cartwright and Taylor (1971) and Cartwright and Edden (1973)
	units	m
	valid_min	-1
	valid_max	1
	tag_basic_expert	Expert
	comment	Solid-Earth (Body) tide height, averaged over the lake. The zero-frequency permanent tide component is not included.
load_tidef		
	type	float
	fill_value	-999999999999
	long_name	geocentric load tide height (FES)
	source	FES2014b (Carrere et al., 2016)
	institution	LEGOS/CNES
	units	m
	valid_min	-0.2
	valid_max	0.2
	tag_basic_expert	Expert
	comment	Geocentric load tide height. The effect of the ocean tide loading of the Earth's crust. This value is used to compute wse.
load_tideg		
	type	float
	fill_value	-999999999999
	long_name	geocentric load tide height (GOT)
	source	GOT4.10c (Ray, 2013)
	institution	GSFC
	units	m
	valid_min	-0.2
	valid_max	0.2
	tag_basic_expert	Expert

	comment	Geocentric load tide height. The effect of the ocean tide loading of the Earth's crust.
pole_tide		
	type	float
	fill_value	-999999999999
	long_name	height of pole tide
	units	m
	source	Wahr (1985) and Desai et al. (2015)
	valid_min	-0.2
	valid_max	0.2
	tag_basic_expert	Expert
	comment	Geocentric pole tide height. The sum total of the contribution from the solid-Earth (body) pole tide height and the load pole tide height (i.e., the effect of the ocean pole tide loading of the Earth's crust).
Geophysical Range Corrections		
dry_trop_c		
	type	float
	fill_value	-999999999999
	long_name	dry tropospheric vertical correction to WSE
	source	European Centre for Medium-Range Weather Forecasting
	institution	ECMWF
	units	m
	valid_min	-3.0
	valid_max	-1.5
	tag_basic_expert	Expert
	comment	Equivalent vertical correction due to dry troposphere delay. Adding the reported correction to the reported lake WSE results in the uncorrected lake WSE.
wet_trop_c		
	type	float
	fill_value	-999999999999
	long_name	wet tropospheric vertical correction to WSE
	source	European Centre for Medium-Range Weather Forecasting
	institution	ECMWF
	units	m
	valid_min	-1
	valid_max	0
	tag_basic_expert	Expert
	comment	Equivalent vertical correction due to wet troposphere delay. Adding the reported correction to the reported lake WSE results in the uncorrected lake WSE.
iono_c		
	type	float
	fill_value	-999999999999
	long_name	ionospheric vertical correction to WSE
	source	Global Ionosphere Maps
	institution	JPL
	units	m
	valid_min	-0.5
	valid_max	0
	tag_basic_expert	Expert
	comment	Equivalent vertical correction due to ionosphere delay. Adding the reported correction to the reported lake WSE results in the uncorrected lake WSE.
Instrument Corrections		
xovr_cal_c		

	type	float
	fill_value	-999999999999
	long_name	crossover calibration height correction
	units	m
	valid_min	-10
	valid_max	10
	tag_basic_expert	Expert
	comment	Equivalent height correction estimated from KaRIn crossover calibration. The correction is applied during processing before geolocation in terms of roll, baseline dilation, etc., but reported as an equivalent height correction. The correction term should be subtracted from the reported WSE to obtain the uncorrected WSE.
Prior Lake Database (PLD) Information		
p_name		
	type	text
	fill_value	“no_data”
	long_name	name(s) of the lake
	comment	Name(s) of the lake, retrieved from Open Street Map, IGN Carthage, GLWD and vMap0 databases. The different names are separated by semicolons.
p_grand_id		
	type	int9
	fill_value	-999999999999
	long_name	reservoir Id from GRanD database
	source	https://doi.org/10.1890/100125
	valid_min	0
	valid_max	10000
	tag_basic_expert	Expert
	comment	Reservoir ID from the Global Reservoir and Dam (GRanD) database. 0=The lake is not a registered reservoir.
p_max_wse		
	type	float
	fill_value	-999999999999
	long_name	maximum water surface elevation
	units	m
	valid_min	-1000
	valid_max	100000
	tag_basic_expert	Basic
	comment	Maximum water surface elevation (except flooding events) from the prior lake database, used to compute storage change.
p_max_area		
	type	float
	fill_value	-999999999999
	long_name	maximum water surface area
	units	km^2
	valid_min	0
	valid_max	500000
	tag_basic_expert	Basic
	comment	Maximum water surface area (except flooding events)from the prior lake database, used to compute storage change.
p_ref_date		
	type	text
	fill_value	“no_data”
	long_name	reference date for the storage change attributes
	tag_basic_expert	Basic

	comment	Reference date from the prior lake database for the storage change attributes, corresponding to the date of the first valid measurement. The format is YYYY-MM-DD.
p_ref_ds		
	type	float
	fill_value	-999999999999
	long_name	Reference storage change
	units	km^3
	valid_min	-1000
	valid_max	1000
	tag_basic_expert	Basic
	comment	Reference storage change from the prior lake database used to translate the storage change values initially computed with respect to the p_max_wse and p_max_area of the PLD lake, to the storage change relative to p_ref_date.
p_storage		
	type	float
	fill_value	-999999999999
	long_name	maximum water storage
	units	km^3
	valid_min	0
	valid_max	30000
	tag_basic_expert	Basic
	comment	Maximum water storage value from the prior lake database, computed between the minimum (or ground when a bathymetry is available) and maximum observed levels of the lake.

5.3 PLD-oriented lake file attribute description

The attributes of the PLD-oriented lake shapefile are the same as those of the observation-oriented lake product described in section 5.2 (Table 9), except for the inversion of the order and roles of *obs_id* and *lake_id,* and the fact that *overlap* is computed the other way around, as described in Table 10. Note that all measurement attributes in this shapefile are aggregated over PLD lakes rather than observed lakes. For records corresponding to unobserved PLD lakes covered by the granule, the attributes are set to the fill values (except for those containing PLD information), and they have no polygon (empty shape).

Table 10. Attributes of the PLD-oriented lake shapefile of the L2_HR_LakeSP product that are MODIFIED with respect to the same attributes in the observation-oriented lake shapefile.

Lake ID		
lake_id		
	type	text
	long_name	lake ID from prior database
	tag_basic_expert	Basic
	comment	Identifier of prior lake from the prior lake database. The format of the identifier is CBBNNNNNNT, where C=continent code, B=basin code, N=lake counter within the basin, T=type. The different lake identifiers are separated by semicolons.
obs_id		
	type	text
	fill_value	“no_data”
	long_name	Identifier(s) of the observed lake(s)
	tag_basic_expert	Basic

	comment	List of identifiers of observed lakes that intersect the prior lake given by lake_id. Unique observation identifier within the product. The format of the identifier is CBBTTTSNNNNNN, where C=continent code, B=basin code, TTT=tile number within the pass, S=swath side, N=lake counter within the tile.
overlap		
	type	text
	fill_value	"no_data"
	long_name	fraction of prior lake covered by each observed lake
	tag_basic_expert	Basic
	comment	List of fractions of prior lake area covered by each observed lake identified in obs_id attribute. The different fractions are separated by semicolons and refer one-to-one to the identifiers listed in the obs_id attribute.

5.4 Observation-oriented unassigned features file attribute description

The attributes of the observation-oriented unassigned features shapefile are the same as those of the observation-oriented lake product described in section 5.2, except that the following attributes listed in Table 9 are NOT included:

- Lake ID: *lake_id, overlap*
- Storage change: *delta_s_l, ds_l_u, delta_s_q, ds_q_u*
- Prior Lake Database (PLD) Information: *p_name, p_grand_id, p_max_wse, p_max_area , p_ref_date, p_ref_ds, p_storage*

6 References

[1] T. Pavelsky and E. H. Altenau, "SWOT L2_HR Prior River Database Auxiliary File Description [TBD]," JPL, 2019.

[2] F. J. Turk and C. Pottier, "Product Description Document - Level 2 KaRIn High Rate River Single Pass Vector Product (L2_HR_RiverSP) - JPL D-56413," JPL, 2019.

[3] C. Pottier and C. Cazals, "SWOT L2 HR Prior Lake Database Auxiliary File Description," CNES, 2019.

[4] Esri, "Esri Shapefile Technical Description - http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf . Brief description available from https://www.loc.gov/preservation/digital/formats/fdd/fdd000280.shtml," Esri White Paper, July 1998.

[5] C. Pottier and F. J. Turk, "Algorithm Theoretical Basis Document (ATBD) - Level 2 KaRIn High Rate Lake Single Pass Science Algorithm Software (SAS_L2_HR_LakeSP)," CNES, 2019.

[6] B. A. Williams and R. Fjortoft, "Product Description Document - Level 2 KaRIn High Rate Water Mask Pixel Cloud Product (L2_HR_PIXC)," JPL, 2019.

[7] J. Furnans and F. Olivera, "Water Shed Topology - The Pfafstetter System," Esri.

[8] B. Lehner and G. Grill, "Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Data is available at www.hydrosheds.org.," Hydrological Processes, 27(15): 2171–2186, 2013.

[9] T. Pavelsky, "SWOT Ice Flag Description," JPL, 2019.

[10] N. K. Pavlis, S. A. Holmes, S. C. Kenyon and J. K. Factor, "The development and evaluation of the Earth Gravitational Model 2008 (EGM2008)," *J. Geophys. Res.: Solid Earth,* vol. 117, pp. 1978-2012, 2012, https://doi.org/10.1029/2011JB008916.

[11] D. E. Cartwright and R. J. Taylor, "New computations of the tide-generating potential," *Geophys. J. R. Astr. Soc.,* vol. 23, pp. 45-74, 1971.

[12] D. E. Cartwright and A. C. Edden, "Corrected tables of tidal harmonics," *Geophys. J. R. Astr. Soc.,* vol. 33, pp. 253-264, 1973.

[13] J. M. Wahr, "Deformation induced by polar motion," *J. Geophys. Res.,* vol. 90(B11), pp. 9363-9368, 1985, https://doi.org/10.1029/JB090iB11p09363.

[14] S. Desai, J. Wahr and B. Beckley, "Revisiting the pole tide for and from satellite altimetry," *J. Geod,* vol. 89, pp. 1233-1243, 2015, https://doi.org/10.1007/s00190-015-0848-7.

[15] J. C. Ries and S. D. Desai, "Conventional model update for rotational deformation," in *Fall AGU Meeting*, New Orleans, LA, 2017, http://dx.doi.org/10.26153/tsw/2659.

[16] L. Carrere, F. Lyard, M. Cancet, A. Guillot and N. Picot, "FES 2014, a new tidal model - Validation results and perpectives for improvements," ESA Living Planet Conference, Prague, 2016.

[17] R. D. Ray, "Precise comparisons of bottom-pressure and altimetric ocean tides," *J. Geophys. Res: Oceans,* vol. 118, pp. 4570-4584, 2013.

[18] N. K. Pavlis, S. A. Holmes, S. C. Kenyon and J. K. Factor, "The development and evaluation of the Earth Gravitational Model 2008 (EGM2008)," *J. Geophys. Res.: Solid Earth,* vol. 117, pp. 1978-2012, 2012, https://doi.org/10.1029/2011JB008916.

[19] D. E. Cartwright and R. J. Taylor, "New computations of the tide-generating potential," *Geophys. J. R. Astr. Soc.,* vol. 23, pp. 45-74, 1971.

[20] D. E. Cartwright and A. C. Edden, "Corrected tables of tidal harmonics," *Geophys. J. R. Astr. Soc.,* vol. 33, pp. 253-264, 1973.

[21] L. Carrere, F. Lyard, M. Cancet, A. Guillot and N. Picot, "FES 2014, a new tidal model - Validation results and perspectives for improvements," in *ESA Living Planet Conference*, Prague, 2016.

[22] R. D. Ray, "Precise comparisons of bottom-pressure and altimetric ocean tides," *J. Geophys. Res: Oceans,* vol. 118, pp. 4570-4584, 2013.

[23] B. Lehner, C. R. Liermann, C. Revenga, C. Vörösmarty, B. Fekete, P. Crouzet, P. Döll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J. C. Robertson, R. Rödel, N. Sindorf and D. Wisser, "High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management," *Frontiers in Ecology and the Environment,* vol. 9, pp. 494-502, 2011.

Appendix A. Acronyms

ATBD	Algorithm Theoretical Basis Document
CNES	Centre National d'Études Spatiales
CRID	Composite Release Identifier
ECMWF	European Center for Medium-Range Weather Forecasts
GSFC	Goddard Space Flight Center
HR	High Rate
JPL	Jet Propulsion Laboratory
KaRIn	Ka-band Radar Interferometer
LR	Low Rate
L2	Level 2
PIXC	Pixel Cloud
PLD	Prior Lake Database
PRD	Prior River Database
RD	Reference Document
SAS	Science Algorithm Software
SDS	Science Data System
SP	Single Pass
SWOT	Surface Water Ocean Topography
TAI	International Atomic Time
TBC	To Be Confirmed
TBD	To Be Determined
UTC	Coordinated Universal Time
XML	Extensible Markup Language

WSE	Water Surface Elevation

Appendix B. **Description of XML files**

The content and the format of the .shp.xml file for the three shapefiles is currently under construction. All content in this section is **TBD**.